National Aeronautics and Space Administration
Goddard Space Flight Center
Contract No. NAS-5-9299

ST - MAT - 10491

ON THE APPLICATION OF DIFFERENCE METHODS FOR THE ASYMPTOTIC ESTIMATES OF ERRORS AT NUMERICAL INTEGRATION OF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

BY
M. L. Brodskiy
(USSR)

24 MAY 1966

ON THE APPLICATION OF DIFFERENCE METHODS FOR THE ASYMPTOTIC ESTIMATES OF ERRORS AT NUMERICAL INTEGRATION OF THE SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS*

Doklady A.N. SSSR,
Matematika,
Tom 43, No. 4, 599 - 602,
Izdatel'stvo AN SSSR, 1953

by M.L. Brodskiy

SUMMARY

The demonstration is based upon a lemma, a theorem and the ensuing corollary. Three particular cases are considered — the Adams, Milne and the Simpson methods. The final formula, arrived at, was tested on a few examples showing good agreement with the computation after introducing an "artificial" error.

* * *

Let us consider a system of r ordinary differential equations written as a single differential equation of the vector of the r-dimensional space:

$$\frac{dy}{dx} = f(x, y) \tag{1}$$

(for the initial condition $y(x_0) = y_0$).

We shall examine the case when the equation (1) is resolved approximately by a difference method of the type

$$y_n - \sum_{j=1}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j f_{n-j}. \tag{2}$$

Let us assume that the approximate solution of the equation (1) is searched for on the segment $[x_0, X]$; let y_n be the approximate value of the solution at the point $x_n = x_0 + nh$; $y(x_n)$ is the exact solution at the same point $\bar{\Delta}_n = y_n - y(x_n)$; $A(x)$ is the matrix $\left\| \frac{\partial f^{(p)}}{\partial y^{(q)}} \right\|_{p, q=1}^{r}$ at the point

---

* ASIMPTOTICHESKIYE OTSENKI POGRESHNOSTEY PRI CHISLENNOM INTEGRIROVANII SISTEM OBYKNOVENNYKH DIFFERENTSIAL'NYKH URAVNENIY RAZNOSTNYMI METODAMI.

$(x, \mathbf{y}(x))$; A_n is the same matrix with values computed at certain intermediate points, so that $\mathbf{f}(x_n, \mathbf{y}_n) - \mathbf{f}(x_n, \mathbf{y}(x_n)) = A_n \bar{\Delta}_n$; $\bar{\rho}_n$ is the error of the method, i.e. the error at substitution of the exact solution $\mathbf{y}(x_n)$ into formula (2); $\bar{\eta}_n$ is the computation error by formula (2) (including the round off error); $\bar{\delta}_n = \bar{\eta}_n - \bar{\rho}_n$; $\bar{\Delta}_0, \bar{\Delta}_1, \ldots, \bar{\Delta}_{k-1}$ are the initial errors, that is, the values of $\mathbf{y}_n - \mathbf{y}(x_n)$ in the first k singular points.

We shall also assume that all the roots of the characteristic equation

$$\lambda^k - \sum_{j=1}^{k} \alpha_j \lambda^{k-j} = 0 \tag{3}$$

do not exceed the unity by module and all of them may be simple besides perhaps zero. We shall seek the expression for $\bar{\Delta}_m$ $(m \geqslant k)$ by $\bar{\Delta}_0, \ldots, \bar{\Delta}_{k-1}, \bar{\delta}_k, \ldots, \bar{\delta}_m$.

Evidently, $\bar{\Delta}_n$ satisfies the difference equation

$$\bar{\Delta}_n - \sum_{j=1}^{k} \alpha_j \bar{\Delta}_{n-j} - h \sum_{j=0}^{k} \beta_j A_{n-j} \bar{\Delta}_{n-j} = \bar{\delta}_n. \tag{4}$$

We shall denote $B(x) = A^*(x_m - x)$; $B_n = A^*_{m-n}$. and construct the vectors $\mathbf{z}_n^{(t)}$ $(t = 1, 2, \ldots, r;\ n = 0, 1, \ldots, m-k)$, satisfying the homogenous difference equation

$$\mathbf{z}_n^{(t)} - \sum_{j=1}^{k} \alpha_j \mathbf{z}_{n-j}^{(t)} - hB_n \sum_{j=0}^{k} \beta_j \mathbf{z}_{n-j}^{(t)} = 0 \tag{5}$$

at initial conditions

$$\mathbf{z}_0^{(t)} - h\beta_0 B_0 \mathbf{z}_0^{(t)} = \mathbf{e}^{(t)} \tag{6}$$

($\mathbf{e}^{(t)}$ are orts of the r-dimensional space),

$$\mathbf{z}_n^{(t)} - \sum_{j=1}^{n} \alpha_j \mathbf{z}_{n-j}^{(t)} - hB_n \sum_{j=0}^{n} \beta_j \mathbf{z}_{n-j}^{(t)} = 0 \quad (1 \leqslant n \leqslant k-1). \tag{7}$$

Then we shall obtain the formula:

$$(\bar{\Delta}_m, \mathbf{e}^{(t)}) = \sum_{n=0}^{m-k} (\bar{\delta}_{m-n}, \mathbf{z}_n^{(t)}) +$$

$$+ \sum_{n=m-k+1}^{m} \left(\bar{\Delta}_{m-n}, \sum_{j=n-m+k}^{k} \alpha_j \mathbf{z}_{n-j}^{(t)} + hB_n \sum_{j=n-m+k}^{k} \beta_j \mathbf{z}_{n-j}^{(t)} \right), \tag{8}$$

giving all the projections of the vector of $\bar{\Delta}_m$ searched for.

Assuming $|\bar{\Delta}_i| = O(h)$ $(0 \leqslant i < k)$, $|\bar{\delta}_n| = O(h^2)$, we shall have by the strength of the condition superimposed on the roots of the equation (3), (see [1]): $|\bar{\Delta}_m| = O(h)$ $(x_0 \leqslant x_m \leqslant X)$, and, assuming the existence and the continuity on the

segment $[x_0, X]$ of second partial derivatives of the type $\frac{\partial^2 f^{(p)}}{\partial y^{(q)} \partial y^{(s)}}$, we shall obtain

$$\| B_n - B(nh) \| = O(h) \tag{9}$$

(in the particular case of linear system (9) is trivially fulfilled even without these assumptions).

At observance of the inequality (9) it may be shown that the solution of the difference equation (5) may be approximated with the help of solutions of certain differential equations. In reality, there takes place the lemma:

LEMMA. - If λ_p is one of roots of the equation (3), different from zero,

$$\sigma_p \equiv \sigma(\lambda_p) = \frac{\sum_{j=0}^{k} \beta_j \lambda_p^{k-j}}{\sum_{j=1}^{k} j \alpha_j \lambda_p^{k-j}}, \tag{10}$$

$\mathbf{u}(x)$ is the solution of the differential equation

$$\frac{d\mathbf{u}}{dx} = \sigma_p B(x)\, \mathbf{u}, \tag{11}$$

$$\mathbf{u}_n = \lambda_p^n \mathbf{u}(nh) \tag{12}$$

and (9) takes place, then we have

$$\left| \mathbf{u}_n - \sum_{j=1}^{k} \alpha_j \mathbf{u}_{n-j} - hB_n \sum_{j=}^{k} \beta_j \mathbf{u}_{n-j} \right| = O(h^2). \tag{13}$$

Resting on this lemma, we may demonstrate the following theorem:

THEOREM.- At initial conditions (6) and (7), the solution of the equation (5) is expressed by the formula

$$z_n^{(q)} = \sum_{p=1}^{l} \tau_p \lambda_p^n \mathbf{u}^{(p,\,q)}(nh) + O(h), \tag{14}$$

where

$$\tau_p = \frac{\lambda_p^k}{\sum_{j=1}^{k} j \alpha_j \lambda_p^{k-j}}; \tag{15}$$

is the number of roots of the equation (3) different from zero; $\mathbf{u}^{(p,\,q)}(x)$ is the solution of the equation (11) at the initial condition $\mathbf{u}(0) = \mathbf{e}^{(q)}$.

DEMONSTRATION.- The solution of the difference equation (5) at initial conditions (6) - (7) amounts to the solution of the same equation

at the condition (6) and $z_{-1}^{(l)} = \ldots = z_{-(k-1)}^{(l)} = 0$.

Expanding in the l-th space the vector $(0, 0, \ldots, 1)$ by vectors $(\lambda_p^{-(l-1)}, \ldots, \lambda_p^{-1}, 1)(p = 1, 2, \ldots, l)$, we shall reduce the solution searched for to the sum of the solutions $z_n^{(p,\,l)}$ with initial conditions $(\lambda_p^{-(k-1)}e^{(l)}, \ldots, e^{(l)})$ with coefficients τ_p and the solution for which the values in the zero, $-1-k, \ldots -(l-1)$-th singular points are zero and which, because of that is equal to $O(h)$ at $n \geqslant 0$.

Examining the difference $v_n^{(p,\,l)} = z_n^{(p,\,l)} - \lambda_p^n u_p^{(l)}(nh)$, we obtain that the values of $v_n^{(p,\,l)}$ at $n = 0, -1, \ldots, -(l-1)$ have the order $O(h)$, so that (13) is valid for $v_n^{(p,\,l)}$, whence $v_n^{(p,\,l)} = O(h)$ over the entire considered segment. Thus the theorem is demonstrated.

COROLLARY.- Under the considered conditions

$$\bar{\Delta}_m = \sum_{l=0}^{k-1}\sum_{p=1}^{l}\tau_{pl}\lambda_p^{m-l}s_p(\bar{\Delta}_l, k, m) + \sum_{n=k}^{m}\sum_{p=1}^{l}\tau_p\lambda_p^{m-k}s_p(\bar{\delta}_n, n, m) + O(h^2), \tag{16}$$

where

$$\tau_{pl} = \frac{\sum_{j=k-l}^{k}\alpha_j\lambda_p^{k-j}}{\sum_{j=1}^{k}j\alpha_j\lambda_p^{k-j}}; \tag{17}$$

$s_p(v, n, m)$ is the solution of the vectorial linear differential equation

$$\frac{dz}{dx} = \sigma_p A(x)z \tag{18}$$

at the point x_m and at initial conditions $z(x_n) = v$.

This corollary stems from the properties of conjugate differential equations.

Let us pause at some particular cases.

1.- For the Adams method (with or without reduction) the only root that is not zero is $\lambda_1 = 1$; $\sigma_1 = \tau_1 = 1$;

$$\bar{\Delta}_m = \sum_{n=k-1}^{m} s_1(\tilde{\delta}_n, n, m) + O(h^2), \tag{19}$$

where $\tilde{\delta}_n = \bar{\delta}_n$ $(n \geqslant k)$; $\tilde{\delta}_n = \bar{\Delta}_n$ $(n < k)$.

2. - For the Milne method

$$y_n - y_{n-4} = h\left(\frac{8}{3} f_{n-1} - \frac{4}{3} f_{n-2} + \frac{8}{3} f_{n-3}\right). \tag{20}$$

The roots of the characteristic equation are $\lambda_1 = 1$, $\lambda_2 = -1$, $\lambda_3 = i$, $\lambda_4 = -i$; $\sigma_1 = 1$, $\sigma_2 = -{}^8/_3$, $\sigma_3 = \sigma_4 = {}^1/_3$; $\tau_1 = \tau_2 = \tau_3 = \tau_4 = {}^1/_4$.

We have :

$$\overline{\Delta}_m = \frac{1}{4} \sum_{n=0}^{m} \sum_{p=1}^{4} s_p(\tilde{\delta}_n, n, m). \tag{21}$$

For the case when the system (1) is reduced to a single equation

$$\frac{dy}{dx} = f(x, y),$$

formula (21) acquires the form :

$$\Delta_m = \frac{1}{4} \sum_{n=0}^{m} \left[\exp\left(\int_{x_n}^{x_m} \frac{\partial f}{\partial y}\, dx \right) + (-1)^{m-n} \exp\left(-\frac{5}{3} \int_{x_n}^{x_m} \frac{\partial f}{\partial y} dx \right) + \right.$$

$$\left. + (i^{m-n} + (-i)^{m-n}) \exp\left(\frac{1}{3} \int_{x_n}^{x_m} \frac{\partial f}{\partial y} dx \right) \right] \tilde{\delta}_n + O(h^2). \tag{22}$$

It may be seen from this formula that the Milne method is unsteady for steady equations and systems [1]. Note that this unsteadiness takes place only for round off errors, since, by the strength of their smoothness the terms of the formula (21), corresponding to $p = 2, 3, 4$, have for the method's errors an order $O(h^2)|\eta|$.

3. - For the Simpson method

$$y_n - y_{n-2} = h({}^1/_3 f_n + {}^4/_3 f_{n-1} + {}^1/_3 f_{n-2}),$$
$$\lambda_1 = 1, \lambda_2 = -1; \sigma_1 = 1, \sigma_2 = -{}^1/_3; \tau_1 = \tau_2 = {}^1/_2, \tag{23}$$

$$\overline{\Delta}_m = \frac{1}{2} \sum_{n=0}^{m} \sum_{p=1}^{2} \lambda_p^{m-n} s_p(\tilde{\delta}_n, n, m) + O(h^2). \tag{24}$$

Note that more precise formulas than, for example (24), may be also obtained from the exact formula (8). For example, assuming

$$|\overline{\Delta}_i| = O(h^3) \ (0 \leqslant i < k), \ |\tilde{\delta}_n| = O(h^3),$$

we have :

$$\overline{\Delta}_m = \frac{1}{2} \sum_{n=0}^{m} [s_1(\tilde{\delta}_n, n-1, m) + (-1)^{m-n} s_2(\tilde{\delta}_n, n-1, m)] + O(h^4). \tag{25}$$

Formula (25) was tested on a few examples : an "artifical" error $\bar{\delta}_n$ was introduced and $\bar{\Delta}_m$ was measured at $m \geqslant n$; the agreement of errors with corrections computed by formula (25) was found to be quite good.

***** THE END *****

Received on
3 July 1953

REFERENCE

[1]. - M. R. SHURA-BURA.- Prikl. matem. i mekhanika., 16, v. 5, 375, 1952.

****

Contract No.NAS-5-9299
Consultants and Designers, Inc.
Arlington, Virginia

Translated by ANDRE L. BRICHANT
on 24 May 1966

DISTRIBUTION

GODDARD SPACE F.C.

100	CLARK, TOWNSEND
110	STROUD
400	BOURDEAU
610	MEREDITH
611	McDONALD (2)
612	HEPPNER
	NESS
613	KUPPERIAN (2)
615	BAUER (2)
614	WHITE (2)
640	H E S S (2)
641	MUSEN
643	SQUIRES SHUTE
540	FLEMING
542	VELEZ • (6)
	CHARNOW
547	SIRY
252	LIBRARY
256	FREAS
630	GI for SS (3)

NASA HQS

SS	NEWELL, NAUGLE
SG	MITCHELL
	SCHARDT
	DUBIN
SL	FELLOWS
	HIPSHER
	HOROWITZ
SM	FOSTER
	GILL
RR	KURZWEG
RRA	WILSON
RTR	NEILL
ATSS	— T
WX	SWEET

OTHER CENTERS

AMES R C

SONETT
LIBRARY

LANGLEY R C

116 KATZOFF
185 WEATHERWAX

JPL

LAWSON sec.314
WYCKOFF
LIBRARY

U.C. BERKELEY

WILCOX

U.IOWA

VAN ALLEN

• requester